CODEC SYSTEM AND METHOD FOR SPATIALLY SCALABLE VIDEO DATA

This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/260,598 filed January 9, 2001.

The present invention relates to a system for coding and decoding video data, and more particularly, to a system and method for coding and decoding spatially scalable video data.

Background of the Invention

Video data is generally processed and transferred in the form of bitstreams. A bitstream is spatially scalable if the bitstream contains at least one independently decodable substream representing video sequence which is less than the full resolution of the bitstream. The conventional standards for processing video data, such as Moving Pictures Experts Group (MPEG) 2, MPEG 4 and H.263+ standards, include spatial scalability modes.

There have been some problems and difficulties in processing video data using the conventional standards. For example, video data processing for spatially scalable video data is inefficient because decoded enhancement-layer sequence has significantly lower video quality than a non-scalable sequence at the same bit rate.

One improved scheme over the conventional video data processing systems and methods is a discrete cosine transform (DCT)-based subband decomposition approach in which subband decomposition is used to create base and enhancement layer data which are then encoded using motion compensated DCT coding. The DCT-based subband decomposition approach is more completely described in "Spatial scalable video coding using a combined

subband-DCT approach", by U. Benzler, Oct. 2000, IEEE, vol. 10, no. 7, pp. 1080-1087.

Although the DCT-based subband decomposition approach provides better efficiency in the spatial scalability mode than the conventional video data processing systems and methods, it still has some disadvantages in processing spatially scalable video data. The disadvantages include the requirement of new quantization matrices for DCT processes. In other words, the DCT-based subband approach creates data whose DCT coefficients are statistically very different from the DCT coefficients of original pixels, and, hence, new quantization matrices for the DCT coefficients must be used to obtain good video quality. The DCT-based subband approach also must be modified to obtain flexibility in varying the relative bit rate allocation between base layer data and enhancement layer data. Thus, an interpolation filter used in the DCT-based subband approach to compute based layer predictions cannot be optimal.

Summary of the Invention

Therefore, a need exists for a video data CODEC system in which DCT-based decimation and interpolation are performed with respect to spatially scalable video data. It would also be advantageous to provide a video data CODEC system that facilitates DCT-based motion compensation scheme for enhancement layer of the video data.

In accordance with the principles of the present invention, there is provided a code-decode (CODEC) system for encoding and decoding spatially scalable video data, including a decimate unit for performing DCT-based down-sampling with respect to macro block data of input video data to produce decimated block data representing the low frequency portion of the macro block data, a first encoder for encoding the decimated block data to produce base layer DCT data having DCT coefficients representing the low frequency portion,

a first decoder for decoding the base layer DCT data from the first encoder to produce base layer block data, an interpolate unit for performing DCT-based interpolation with respect to the base layer block data from the first decoder to produce interpolated base layer block data, a second encoder for encoding enhancement layer block data obtained from the macro block data and the interpolated base layer block data to produce enhancement layer DCT data wherein the enhancement layer block data represents the high frequency portion of the macro block data, and a second decoder for decoding the enhancement layer DCT data from the second encoder to produce reconstructed macro block data.

In an exemplary embodiment of the invention, the first encoder includes a first motion compensate unit for compensating the decimated block data from the decimate unit with base layer data of a previous picture and motion vectors for the macro block data to produce compensated base layer block data so that the first encoder performs DCT with respect to the compensated base layer block data to produce the base layer DCT data.

The second encoder may also include a second motion compensation unit for compensating the enhancement layer block data with enhancement layer data of a previous picture and motion vectors for the macro block data to produce compensated enhancement layer block data so that the second encoder performs DCT with respect to the compensated enhancement layer block data to produce the enhancement layer DCT data.

Brief Description of the Drawings

This disclosure will present in detail the following description of preferred embodiment with reference to the following figures wherein:

Fig. 1 is a block diagram of a CODEC system according to a preferred embodiment of the present invention;

Fig. 2 is a block diagram of the decimate unit in Fig. 1 according to a preferred embodiment of the present invention;

Fig. 3 is a block diagram of the interpolate unit in Fig. 1 according to a preferred embodiment of the present invention;

Fig. 4 is a more detailed block diagram of the CODEC system in Fig. 1;

Fig. 5 is a block diagram of the first motion compensation unit in Fig. 4 according to a preferred embodiment of the present invention; and

Fig. 6 is a block diagram of the second motion compensation unit in Fig. 4 according to a preferred embodiment of the present invention.

Detailed Description of Preferred Embodiments

A video data code-decode (CODEC) system according to the present invention employs discrete cosine transform (DCT)-based manipulation (e.g., decimation and interpolation) of video data. DCT-based sampling puts low frequency DCT contents of a block into low resolution pixels of a base layer and leaves pixels containing high frequency DCT contents to be encoded in an enhancement layer. Horizontal and vertical DCT-based decimation and interpolation of a pixel block can be implemented in the spatial domain as a single matrix multiplication, thereby its complexity can be reduced.

In addition to the new DCT-based decimation/interpolation for spatial scalability, the video data CODEC system of the present invention also employs a new motion compensation scheme that predicts pixels containing low-frequency DCT contents in a base layer and pixels containing high-frequency DCT contents in an enhancement layer from corresponding base layer and enhancement layer, respectively, of the previous reconstructed picture. In this approach of the present invention, a single set of motion vectors is used for both the base layer and the enhancement layer.

The video data CODEC system of the present invention applying to spatial scalability modes does not have the disadvantages of the conventional CODEC

systems such as the DCT-based subband approach. Since the DCT-based sample-rate conversion in the present invention is matched to DCT contents used for the subsequent coding of pixels or prediction errors, good performance can be obtained using the same quantization matrices as those used for a non-scalable CODEC. Thus, the quantization matrices can be optimized. The video data CODEC system with new spatial scalability algorithm is described in detail below.

Referring to Fig. 1, there is provided a block diagram for illustrating the creation of the base and enhancement layer data for a macro block of input video data using DCT-based decimation and interpolation. A CODEC system of the present invention includes a decimate unit 11, a first encoder 13, a first decoder 15, an interpolate unit 17, a second encoder 18, and a second decoder 19. The CODEC system receives video data in the form of frames each of which has a predetermined number of blocks. Each block may have a predetermined size, that is, a predetermined number of pixels. For the purpose of the description, it is assumed that video data is applied to the CODEC system in the form of 16x16 blocks in sequence. A 16x16 macro block data B_{16x16} applied to the CODEC system is inputted into the decimate unit 11. The decimate unit 11 performs DCT-based decimation with respect to the input macro block B_{16x16}. Preferably, the decimation is independently performed on each of the four 8x8 blocks in the 16x16 macro block so that the macro block is converted into an 8x8 block.

The first encoder 13 encodes the decimated block data output from the decimate unit to obtain base layer data containing low frequency contents of the macro block B_{16x16}. The first encoder 13 also performs motion compensation with respect to the block data from the decimate unit 11 using base layer data of the previous picture. The first decoder 15 decodes the encoded data from the first encoder 11 to produce 8x8 block data.

The interpolate unit 17 performs DCT-based interpolation with respect to the 8x8 block data received from the first decoder 15 to produce a 16x16 macro block. The original macro block data B_{16x16} is subtracted by the macro block data output from the interpolate unit 17 in the adder.

The second encoder 18 encodes the output data of the adder 16 to obtain enhancement layer data containing high frequency contents of the input macro block data B_{16x16}. In the second encoder 18, motion compensation is performed with respect to the output data of the adder 16 using enhancement layer data of the previous reconstructed picture. The second decoder 19 decodes the encoded data received from the second encoder 18 to produce reconstructed 16x16 macro block data. Each block of the CODEC system in Fig. 1 is described in detail below.

Referring to Fig. 2, there is provided a block diagram of the decimate unit 11 in Fig. 1 according to a preferred embodiment of the present invention. Since the decimate unit 11 performs DCT-based decimation with respect to each of the four 8x8 blocks in the macro block, an 8x8 DCT unit 21 receives an 8x8 block data B_{8x8} and transforms the block into 8x8 DCT coefficients to be provided to a truncate unit 23. In the truncate unit 23, all of the 8x8 DCT coefficients except for the low frequency coefficients, for example, the upper left 4x4 DCT content, are discarded. The remaining low frequency coefficients are provided to a 4x4 inverse DCT (IDCT) unit 25 and transformed into a 4x4 block data B_{4x4}. This series of decimation or down-sampling operations can be represented by the equation (1).

$$B_{4x4} = M_1 B_{8x8} M_2 \qquad (1)$$

Here, M_1 is a 4x8 matrix and M_2 is an 8x4 matrix. To obtain the base layer 4x4 block B_{4x4}, the 8x8 block B_{8x8} is vertically down-sampled with matrix M_1 and horizontally down-down-sampled with matrix M_2.

Referring to Fig. 3, there is provided a block diagram of the interpolate unit 17 in Fig. 1 according to a preferred embodiment of the present invention. The interpolate unit 17 performs DCT-based interpolation with respect to each of four 4x4 blocks in an input 8x8 block. A 4x4 DCT unit 31 receives a 4x4 block data B_{4x4} to transform the block B_{4x4} into 4x4 DCT coefficients. In a zero pad unit 33, the 4x4 block of DCT coefficients is padded with zeros to form an 8x8 block of DCT coefficients. An 8x8 IDCT unit 35 receives the 8x8 DCT coefficients from the zero pad unit 33 and produce an interpolated 8x8 block data. This series of interpolation or up-sampling operations can be represented by the equation (2).

$$B_{8x8} = H_1 B_{4x4} H_2 \qquad (2)$$

Here, H_1 is an 8x4 matrix and H_2 is a 4x8 matrix. To obtain the interpolated 8x8 block B_{8x8}, the 4x4 block B_{4x4} is vertically down-sampled with matrix H_1 and horizontally down-sampled with matrix H_2.

Referring to Fig. 4, there is provided a block diagram for illustrating in detail the first and second encoders 13, 18 and the first and second decoders 15, 19 in Fig. 1. It should be noted that the detailed block diagram in Fig. 4 shows an exemplary embodiment of the present invention for the purpose of description. The first encoder 13 includes a first motion compensation unit 131 for compensating the decimated block data from the decimate unit 11 and a DCT and quantization unit 133 for transforming a 8x8 block of the compensated base layer data into 8x8 DCT coefficients which is then quantized. The quantized DCT coefficients are preferably entropy coded (EC) to produce base

layer bitstream. The base layer bitstream will be reconstructed for the use of motion compensation of the next picture.

Referring to Fig. 5, there is provided a block diagram of the first motion compensation unit 131 in Fig. 4 according to a preferred embodiment of the present invention. In the first motion compensation unit 131, a select unit 51 receives base layer data of the previous reconstructed picture and motion vectors for the 16x16 macro block data to select a block of pixels to be predicted from the previous reconstructed base layer data. It is assumed that motion estimation has been performed at the full resolution. Preferably, one motion vector is used for each 8x8 block This block of pixels is obtained after dividing the motion vector for the block in half, since the motion estimation has been performed at full resolution and the base layer frames have only half as many pixels in both the horizontal and vertical directions as the full-resolution frames.

An interpolate unit 53 performs DCT-based interpolation with respect to the block of pixels selected by the select unit 51 to produce a prediction block with full resolution. The full resolution prediction block may be subject to half-pixel interpolation 55, if necessary, using simple bilinear interpolation. The motion compensation will be done with half-pixel accuracy by use of the half-pixel interpolation. In other words, a motion vector may point to a location halfway between two pixels in a reference frame, in which case the prediction is computed using interpolation. The need to do half-pixel interpolation is indicated by an odd value for the motion vector. A decimate unit 57 performs DCT-based decimation to down-sample the full resolution prediction block from the interpolate block or the half-pixel interpolated block to produce a base layer prediction block.

Referring again to Fig. 4, the base layer prediction block generated from the first motion compensation unit 131 is subtracted from the decimated block

data to obtain the compensated base layer block data. The base layer prediction block is also provided to the first decoder 15 and added to an output of an IDCT and inverse quantization unit 151 which performs IDCT and inverse quantization (IQ) with respect to the output of the DCT and quantization unit 133 in the first encoder 13. In this embodiment, the DCT, IDCT, quantization, IQ, and entropy coding (EC) are well known in the art, thus a detailed description thereof is omitted.

Referring to Fig. 6, there is provided a block diagram of a second motion compensation unit 181 in the second encoder 18 according to a preferred embodiment of the present invention. In the second motion compensation unit 181, a select unit 61 receives enhancement layer data of the previous reconstructed picture and motion vectors for the 16x16 macro block data to select a block of pixels to be predicted from enhancement layer data of the previous reconstructed picture. A DCT unit 63 converts the block of pixels to be predicted into 8x8 DCT coefficients. Upon receiving the DCT coefficients, a low frequency remove unit 65 removes low frequency contents of the 8x8 DCT coefficients by setting low frequency DCT coefficients to zero. As a result, the block of pixels to be predicted for the enhancement layer motion compensation contains only high frequency DCT coefficients. An IDCT unit 67 converts the high frequency DCT coefficients into an 8x8 block data.

Such enhancement layer motion compensation process is preferably implemented as a matrix multiplication for each of the horizontal and vertical directions. In other words, to implement the enhancement layer motion compensation, the operations shown in the DCT unit 63, the low frequency remove unit 65, and the IDCT unit 67 can be combined into a pre-multiplication of the DCT unit 63 input by one matrix for vertical processing and a post-multiplication by another matrix for horizontal processing. This is much the same as the implementation of the decimate unit 11 and the interpolate unit 17

in Figs. 2 and 3, respectively, using matrix multiplication as shown in Equations (1) and (2).

Referring again to Fig. 4, block data output from the second motion compensation unit 181 is subtracted from the block data obtained by subtracting the output of the interpolate unit 17 from the input macro block data B_{16x16}. As a result, compensated enhancement layer block data is obtained. In DCT and quantization units 183, each of four 8x8 blocks of the compensated enhancement layer block data (16x16) is converted into 8x8 DCT coefficients which are then quantized to produce enhancement layer DCT data. The second decoder 18 generates enhancement layer bitstream by entropy coding the enhancement layer DCT data from each of the DCT and quantization units 183.

The second decoder 19 has IQ and IDCT units 191 each for performing inverse quantization and inverse DCT with respect to the enhancement layer DCT data provided from a corresponding one of the DCT and quantization units 183 in the second encoder 18. The IQ and IDCT units 191 produce enhancement layer block data (i.e., 16x16 macro block) by combining four 8x8 blocks each generated from the respective IQ and IDCT units 183. The enhancement layer block data is added to the output of the second motion compensation unit 181, and a result thereof is also added to the output of the interpolate unit 17 to produce a reconstructed macro block data. The two additions in the second decoder 19 correspond to the two subtractions which are the subtraction of the output of the interpolate unit 17 from the input macro block data B_{16x16} and the subtraction of the output of the second motion compensation unit 181 from the difference obtained by the previous subtraction.

In Fig. 4, the first and second encoders 13, 18 have the DCT and quantization units 133, 183 and the first and second decoders 15, 19 have the IQ and IDCT units 151, 191. It should be noted that the DCT and quantization functions or IQ and IDCT functions are combined into one unit only for the

purpose of simplifying the embodiment in Fig. 4. The quantization, IQ, DCT, and IDCT may be implemented in separate units respectively.

In a preferred embodiment of the present invention, there is provided a method for coding and decoding spatially scalable video data, including the steps of decimating macro block data of input video data by performing DCT-based down-sampling to obtain decimated block data representing low frequency part of the macro block data, encoding the decimated block data to obtain base layer DCT data having DCT coefficients representing the low frequency part, decoding the base layer DCT data to obtain base layer block data, interpolating the base layer block data by performing DCT-based up-sampling with respect to the base layer block data to produce interpolated base layer block data, subtracting the interpolated base layer block data from the macro block data to obtain enhancement layer block data representing high frequency part of the macro block data, encoding the enhancement layer block data to obtain enhancement layer DCT data, and decoding the enhancement layer DCT data to obtain reconstructed macro block data.

The encoding the decimated block data step may include performing motion compensation of the decimated block data, wherein the motion compensation includes selecting a block of pixels to be predicted from base layer data of a previous picture using motion vectors for the macro block data, performing DCT-based interpolation with respect to the block of pixels to be predicted to produce full resolution prediction block data, decimating the full resolution prediction block data by performing DCT-based down-sampling to produce base layer prediction block data, and subtracting the base layer prediction block data from the decimated block data to obtain compensated base layer block data which is subject to DCT to produce the base layer DCT data.

The encoding the enhancement layer block data step may also include performing motion compensation of the enhancement layer block data, wherein

the motion compensation includes selecting a block of pixels to be predicted from enhancement layer data of a previous picture using motion vectors for the macro block data, performing DCT with respect to the block of pixels to be predicted to produce a block of DCT coefficients, removing low frequency contents of the block of DCT coefficients, performing IDCT with respect to the DCT coefficients in which the low frequency contents are removed, to obtain enhancement layer prediction block data, and subtracting the enhancement layer prediction block data from the enhancement layer block data to obtain compensated enhancement layer block data which is subject to DCT to produce the enhancement layer DCT data.

Having described preferred embodiments of a CODEC system for spatially scalable video data according to the present invention, modifications and variations can be readily made by those skilled in the art in light of the above teachings. It is therefore to be understood that, within the scope of the appended claims, the present invention can be practiced in a manner other than as specifically described herein.